ENGINEERING β-KETOACYL ACP SYNTHASE FOR NOVEL SUBSTRATE SPECIFICITY

INTRODUCTION

This application claims the benefit of U.S. Provisional Application Number 60/138,308 filed June 9, 1999.

Technical Field

The present invention is directed to proteins, nucleic acid sequences and constructs, and methods related thereto.

Background

Fatty acids are organic acids having a hydrocarbon chain of from about 4 to 24 carbons. Many different kinds of fatty acids are known which differ from each other in chain length, and in the presence, number and position of double bonds. In cells, fatty acids typically exist in covalently bound forms, the carboxyl portion being referred to as a fatty acyl group. The chain length and degree of saturation of these molecules is often depicted by the formula CX:Y, where "X" indicates number of carbons and "Y" indicates number of double bonds.

The production of fatty acids in plants begins in the plastid with the reaction between acetyl-CoA and malonyl-ACP to produce acetoacetyl-ACP catalyzed by the enzyme, ß-ketoacyl-ACP synthase III. Elongation of acetyl-ACP to 16- and 18- carbon fatty acids involves thefollowing cycle of reactions: condensation with a two-carbon unit from malonyl-ACP to form a ß-ketoacyl-ACP (ß-ketoacyl-ACP synthase), reduction of the keto-function to an alcohol (ß-ketoacyl-ACP reductase), dehydration to form an enoyl-ACP (ß-hydroxyacyl-ACP dehydrase), and finally reduction of the enoyl-ACP to form the elongated saturated acyl-ACP (enoyl-ACP reductase). ß-ketoacyl-ACP synthase I, catalyzes elongation up to palmitoyl-ACP (C16:0), whereas ß-ketoacyl-ACP synthase II catalyzes the final elongation to stearoyl-ACP (C18:0). The longest chain fatty acids produced by the

FAS are typically 18 carbons long. Additional biochemical steps in the cell produce specific fatty acid constituents, for example through desaturation and elongation.

β-ketoacyl synthases, condensing enzymes, comprise a structurally and functionally related family that play critical roles in the biosynthesis of a variety of natural products, including fatty acids, and the polyketide precursors leading to antibiotics, toxins, and other secondary metabolites. β-ketoacyl synthases catalyze carbon-carbon bond forming reactions bycondenisng a variety of acyl chain precursors with an elongating carbon source, usually malonyl or methyl malonyl moieties, that are covalently attached through a thioester linkage to an acyl carrier protein. Condensing enzymes can be part of multienzyme complexes, domains of large, multifunctional polypeptide chains as the mammalian fatty acid synthase, or single enzymes as the β-ketoacyl synthases in plants and most bacteria.

Condensing enzymes have been identified with properties subject to exploitation in the areas of plant oil modification, polyketide engineering, and ultimately design anti-cancer and anti-tuberculosis agents. One of the molecular targets of isoniazid, which is widely used in the treatment of tuberculosis, is KAS. Cerulinin, a mycotoxin produced by the fungus *Cephalosporium caerulens*, acts as a potent inhibitor of KAS by covalent modification of the active cysteine thiol. Condensing enzymes from many other pathways and sources have all been shown to be inactivated by this antibiotic with the exception of the synthase from *C. caerulens* and KASIII, the isozyme responsible for the initial condensation of malonyl-ACP with acetyl-CoA in plant and bacterial fatty acid biosynthesis. Inhibition of the KAS domain of fatty acid synthase bycerulinin is selectively cytotoxic to certain cancer cells.

SUMMARY OF THE INVENTION

The present invention is directed to β-ketoacyl ACP synthase (KAS), and in particular to engineered KAS polypeptides and polynucleotides encoding engineered KAS proteins having a modified substrate specificity with respect to the native (also referred to herein as wild-type) KAS protein. The engineered polypeptides and polynucleotides of the present invention include those derived from plant and bacterial sources.

In another aspect of the invention polynucleotides encoding engineered polypeptides, particularly, polynucleotides that encode a KAS protein with a modified substrate specificity with respect to the native KAS protein, are provided.

In a further aspect the invention relates to oligonucleotides derived from the engineered KAS proteins and oligonucleotides which include partial or complete engineered KAS encoding sequences.

It is also an aspect of the present invention to provide recombinant DNA constructs which can be used for transcription or transcription and translation (expression) of an engineered KAS protein having an altered substrate specificity with respect to the native KAS protein. In particular, constructs are provided which are capable of transcription or transcription and translation in host cells. Particularly preferred constructs are those capable of transcription or transcription and translation in plant cells.

In another aspect of the present invention, methods are provided for production of engineered KAS proteins having a modified substrate specificity with respect to the native KAS in a host cell or progeny thereof. In particular, host cells are transformed or transfected with a DNA construct which can be used for transcription or transcription and translation of an engineered KAS. The recombinant cells which contain engineered KAS are also part of the present invention.

In a further aspect, the present invention relates to methods of using the engineered polynucleotide and polypeptide sequences of the present invention to modify the fatty acid composition in a host cell, as well as to modify the composition and/or structure of triglyceride molecules, particularly in seed oil of oilseed crops. Plant cells having such a modified triglyceride content are also contemplated herein.

The modified plants, seeds and oils obtained by the expression of the plant engineered KAS proteins are also considered part of the invention.

DESCRIPTION OF THE FIGURES

Figures 1-1 to 1-48 DIS 01/03/02

~~Figure 1~~ provides the coordinates of the crystal structure of the *E. coli* KAS protein. The first column provides the Type of atom (N=Nitrogen, O=oxygen, C=Carbon, CA= alpha carbon, CB=beta carbon, CG= gamma carbon, CD= delta carbon, CE= epsilon carbon, NZ= zeta

nitrogen, NH= amino group), the second column provides the amino acid residue type (three letter abbreviation), the third column provides the subunit in which the amino acid is located, the forth column provides the amino acid position in the protein sequence base don the mature unprocessed protein, columns seven through nine provide the x, y and z coordinates, respectively, of the three dimensional location of the respective atom in the crystal structure.

DJS 1/3/01 Figures 2-1 to 2-49 ~~Figure 2~~ provides the profile of the crystal structure of the *E. coli* KAS-cerulenin complex. The first column provides the Type of atom (N=Nitrogen, O=oxygen, C=Carbon, CA= alpha carbon, CB=beta carbon, CG= gamma carbon, CD= delta carbon, CE= epsilon carbon, NZ= zeta nitrogen, NH= amino group), the second column provides the amino acid residue type (three letter abbreviation), the third column provides the subunit in which the amino acid is located, the forth column provides the amino acid position in the protein sequence base don the mature unprocessed protein, columns seven through nine provide the x, y and z coordinates, respectively, of the three dimensional location of the respective atom in the crystal structure.

Figure 3 provides the effects of KAS II mutations on the fatty acid composition of *E. coli*.

Figure 4 shows that mutations I108F, I108L and A193M all cause significant reduction in the activity of KAS II on 8:0-ACP as compared to 6:0-ACP (38, 31 and 12 fold reductions respectively), without significantly reducing the activity on 6:0-ACP.

Figure 5 shows that the combined mutations at I108 and A193 have the effect of reducing the activity of KAS II on 6:0-ACP substrates.

Figure 6 shows that the combined effect of two or more mutations had a greater effect on the activity with acyl-ACPs 8:0 and longer (14:0) substrates.

Figure 7 shows the complete list of mutations that were generated.

Figure 8 provides the structure of the *Cpu* KAS I homodimer

Figure 9 provides the structure of the *Cpu* KAS IV homodimer

Figure 10 provides the structure of the *Cpu* KAS I/ *Cpu* KAS IV heterodimer.

Figure 11 provides the sequence differences in the hydrophobic pocket of the *E. coli* KASII and *C. pu* KASIV.

DJS 1/3/01 Figures 12-1 to 12-5 ~~Figure 12~~ provides an amino acid sequence alignment of KAS protein sequences from plant (*Arabidopsis, Brassica, Cuphea hookeriana and pullcherima, Hordeum, Riccinus*), bacterial (*E. coli, streptococcus,* tuberculosis), mammalian (rat, mouse) and others (*C.elegans*).

Figure 13 provides a bar graph representing the results of fatty acid analysis of seeds from transformed *Arabidopsis* lines containing pCGN11058, pCGN11062, pCGN11041, or nontransformed control lines (AT002-44). For each line, bars represent, from left to right, C12:0, C14:0, C16:0, C16:1, C18:0, C18:1 (delta 9), C18:1 (delta 11), C18:2, C18:3, C20:0, C20:1 (delta 11), C20:1 (delta 13), C20:2, C20:3, C22:0, C22:1, C22:2, C22:3, C24:0, and C24:1 fatty acids.

Figure 14 provides the nucleotide sequence of the plastid targeting sequence from *Cuphea hookeriana* KASII-7.

DETAILED DESCRIPTION OF THE INVENTION

In accordance with the subject invention, engineered nucleotide sequences are provided which are capable of coding sequences of amino acids, such as, a protein, polypeptide or peptide. The engineered nucleotide sequences encode β-ketoacyl-ACP synthase (KAS) proteins with a modified substrate specificity compared to the native KAS protein (also referred to herein as the wild-type KAS protein) under enzyme reaction conditions. Such sequences are referred to herein as engineered β-ketoacyl-ACP synthase (also referred to as engineered KAS) proteins. The engineered nucleic acid sequences find use in the preparation of constructs to direct their expression in a host cell. Furthermore, the engineered nucleic acid sequences find use in the preparation of plant expression constructs to alter the fatty acid composition of a plant cell. By "enzyme reactive conditions" is meant that any necessary conditions are available in an environment (for example, such factors as temperature, pH, lack of inhibiting substances) which will permit the enzyme to function.

An engineered β-ketoacyl-ACP synthase nucleic acid sequence of this invention includes any nucleic acid sequence coding a β-ketoacyl-ACP synthase having altered substrate specificity relative to the native KAS in a host cell, includign but not limited to, *in vivo*, or in a cell-like environment, for example, *in vitro*. By altered, or modified, substrate specificity is meant an alteration in the acyl-ACP substrates elongated by the KAS enzyme or an alteration in the elongator molecule used by the KAS to elongate the acyl-ACP relative to the native or unaltered KAS protein. An alteration in the acyl-ACP substrate elongated by the KAS enzymes includes,

but is not limited to, elongation of an acyl-ACP substrate not elongated by the wild-type KAS, the inability to elongate an acyl-ACP substrate elongated by the wild-type KAS, and a preference for elongating acyl-ACP substrates not normally preferred by the wild-type KAS. An alteration in the elongator molecule used by the engineered KAS for the elongation of the acyl-ACP substrate includes, but is not limited to, methyl-malonyl ACP for the production of branched chain fatty acids.

A first aspect of the present invention relates to engineered β-ketoacyl-ACP synthase polypeptides. In particular, engineered KAS II polypeptides are provided. Preferred peptides include those found in the hydrophobic fatty acid/cerulenin binding pocket of the KAS protein. Such polypeptides include the engineered polypeptides set forth in the Sequence Listing, as well as polypeptides and fragments thereof, particularly those polypeptides which exhibit a modified substrate specificity with respect to the wild-type KAS polypeptide. Particularly preferred polypeptides include those having engineered amino acid residues 105 to 120, 130-140, 190-200 and 340-400. Most preferred polypeptides include those having engineered amino acid residues I108A, I108F, I108G, I108L, L111A, I114A, F133A, V134A, V134G, I138A, I138G, A162G, A193G, A193I, A193M, L197A, F202L, F202I, F202G, L342A, and L342G. Amino acid positions, as used herein, refer to the amino acid residue position in the active or processed protein.

Engineered β-ketoacyl-ACP synthases can be prepared by random (via chemical mutagenesis or DNA shuffling) or specific mutagenesis of a β-ketoacyl-ACP synthase encoding sequence to provide for one or more amino acid substitutions in the translated amino acid sequence. Alternatively, an engineered β-ketoacyl-ACP synthase can be prepared by domain swapping between related β-ketoacyl-ACP synthases, wherein extensive regions of the native β-ketoacyl-ACP synthase encoding sequence are replaced with the corresponding region from a different β-ketoacyl-ACP synthase.

Altered substrate specificities of an engineered β-ketoacyl-ACP synthase can be reflected by the elongation of an acyl-ACP substrates of particular chain length fatty acyl-ACP groups which are not elongated by the native β-ketoacyl-ACP synthase enzyme. In addition, altered substrate specificities can be reflected by the in ability to elongate an acyl-ACP substrate of

particular chain length fatty acyl-ACP groups which are not normally preferred by the native β-ketoacyl-ACP synthase enzyme. The newly recognized acyl-ACP substrate can differ from native substrates of the enzyme in various ways, such as by having a shorter or longer carbon chain length (usually reflected by the addition or deletion of one or more 2-carbon units), as well as by degrees of unsaturation.

Another aspect of the present invention relates to engineered β-ketoacyl-ACP synthase polynucleotides. In particular, engineered β-ketoacyl-ACP synthase II polynucleotides are provided. The polynucleotide sequences of the present invention include engineered polynucleotides that encode the polypeptides of the invention having a deduced amino acid sequence selected from the group of sequences set forth in the Sequence Listing.

The invention provides a polynucleotide sequence identical over its entire length to each coding sequence as set forth in the Sequence Listing. The invention also provides the coding sequence for the mature polypeptide or a fragment thereof, as well as the coding sequence for the mature engineered polypeptide or a fragment thereof in a reading frame with other coding sequences, such as those encoding a leader or secretory sequence, a pre-, pro-, or prepro- protein sequence. The polynucleotide can also include non-coding sequences, including for example, but not limited to, non-coding 5' and 3' sequences, such as the transcribed, untranslated sequences, termination signals, ribosome binding sites, sequences that stabilize mRNA, introns, polyadenylation signals, and additional coding sequence that encodes additional amino acids. For example, a marker sequence can be included to facilitate the purification of the fused polypeptide. Polynucleotides of the present invention also include polynucleotides comprising a structural gene and the naturally associated sequences that control gene expression.

As described herein, analysis of the KAS II/cerulinin crystal structure complex is performed using modeling software to produce a profile of the complex, as well as the KAS II protein alone. Based on comparisons of the two profiles, amino acid residues are identified, which when mutagenized, alter the fatty acyl substrate specificities. As demonstrated herein, engineering of the nucleic acid sequence to modify the amino acid sequence in particular regions of the KAS protein effectively modify the substrate specificity of the engineered KAS. Particular ranges for the engineering of the protein include amino acid residues 105 to 120, 130-140, 190-200 and 340-345. Particularly, engineering of residues 108, 111, 114, 133, 193 and 197 can alter

the length of the fatty acids synthesized by the engineered KAS II protein. More particularly, engineering of residues 108, 111, 114, 133, 193 and 197 with variously sized hydrophobic residues will alter the length of the fatty acids synthesized by the engineered KAS II protein. Furthermore, engineering the amino acid residue at position 400 can also have an effect on the substrate specificity.

As demonstrated more fully in the following examples, the acyl-ACP substrate specificity of b-ketoacyl-ACP synthases may be modified by various amino acid changes to the protein sequence, such as amino acid substitutions, insertions or deletions in the mature protein portion of the b-ketoacyl-ACP synthases. Modified substrate specificity can be detected by expression of the engineered b-ketoacyl-ACP synthase s in *E. coli* and assaying to detect enzyme activity or by using purified protein in *in vitro* assays.

Modified substrate specificity can be indicted by a shift in acyl-ACP substrate preference such that the engineered b-ketoacyl-ACP synthase is newly capable of utilizing a substrate not recognized by the native b-ketoacyl-ACP synthase . The newly recognized substrate can vary from substrates of the native enzyme by carbon chain length and/or degree of saturation of the fatty acyl portion of the substrate. Additionally, modified substrate specificity can be reflected by a shift in the relative b-ketoacyl-ACP synthase activity on two or more substrates of the native b-ketoacyl-ACP synthase such that an engineered b-ketoacyl-ACP synthase exhibits a different order of preference for the acyl-ACP substrates.

Furthermore, provided herein are KAS proteins with an altered elongator molecule preference. For example, by widening the hydrophobic fatty acid binding differentelongator molecules, other than Malonyl-ACP, can be utilized by the KAS protein. For example Methyl-malonyl-ACP can be utilized by the engineered KAS resulting in the synthesis of branched chained fatty acid. The mutations that lengthen the pocket may to some degree also widen it, in addition mutations A193G, I108G, L342A or G, V134A or G,F202L,I or G may well cause widening of the pocket sufficiently to allow Methyl-malonyl-ACP to be accepted as an elongator.

As described in more detail herein, alterations in the nucleic acid sequence of the *E. coli* KAS II, particularly, I108F, I108L, A193I, A193M, as well as combinations thereof, are prepared for the production of shorter chain length fatty acids. Furthermore, alterations of I108A, L111A,

I114A, F133A, L197A, and combinations thereof, are prepared for increasing the length of fatty acids produced by the host cell.

Thus, as the result of modifications to the substrate specificity of b-ketoacyl-ACP synthases, it can be seen that the relative amounts of the fatty acids produced in a cell where various substrates are available for hydrolysis may be altered. Furthermore, molecules which are formed from available free fatty acids, such as plant seed triglycerides, may also be altered as a result of expression of engineered b-ketoacyl-ACP synthase s having altered substrate specificities.

It is anticipated that the ranges of mutations provided herein can also be engineered in plant KAS proteins as well as to other polyketide synthases. Such plant KAS proteins are known in the art, and are described for example in PCT Publication WO 98/46776, and in U.S. Patent Number 5,475,099, the entireties of which are incorporated herein by reference.

Plant Constructs and Methods of Use

Of particular interest is the use of the nucleotide sequences, or polynucleotides, in recombinant DNA constructs to direct the transcription or transcription and translation (expression) of the engineered KAS sequences of the present invention in a host plant cell. The expression constructs generally comprise a promoter functional in a host plant cell operably linked to a nucleic acid sequence encoding a engineered KAS of the present invention and a transcriptional termination region functional in a host plant cell.

Those skilled in the art will recognize that there are a number of promoters which are functional in plant cells, and have been described in the literature. Chloroplast and plastid specific promoters, chloroplast or plastid functional promoters, and chloroplast or plastid operable promoters are also envisioned.

One set of promoters are constitutive promoters such as the CaMV35S or FMV35S promoters that yield high levels of expression in most plant organs. Enhanced or duplicated versions of the CaMV35S and FMV35S promoters are useful in the practice of this invention (Odell, *et al.* (1985) *Nature* 313:810-812; Rogers, U.S. Patent Number 5,378, 619). In addition, it may also be preferred to bring about expression of the engineered KAS in specific tissues of the

plant, such as leaf, stem, root, tuber, seed, fruit, etc., and the promoter chosen should have the desired tissue and developmental specificity.

Of particular interest is the expression of the nucleic acid sequences of the present invention from transcription initiation regions which are preferentially expressed in a plant seed tissue. Examples of such seed preferential transcription initiation sequences include those sequences derived from sequences encoding plant storage protein genes or from genes involved in fatty acid biosynthesis in oilseeds. Examples of such promoters include the 5' regulatory regions from such genes as napin (Kridl *et al.*, *Seed Sci. Res. 1*:209:219 (1991)), phaseolin, zein, soybean trypsin inhibitor, ACP, stearoyl-ACP desaturase, soybean α' subunit of β-conglycinin (soy 7s, (Chen *et al.*, *Proc. Natl. Acad. Sci.*, 83:8560-8564 (1986))) and oleosin.

It may be advantageous to direct the localization of proteins to a particular subcellular compartment, for example, to the mitochondrion, endoplasmic reticulum, vacuoles, chloroplast or other plastidic compartment. For example, where the genes of interest of the present invention will be targeted to plastids, such as chloroplasts, for expression, the constructs will also employ the use of sequences to direct the gene to the plastid. Such sequences are referred to herein as chloroplast transit peptides (CTP) or plastid transit peptides (PTP). In this manner, where the protein of interest is not directly inserted into the plastid, the expression construct will additionally contain a gene encoding a transit peptide to direct the protein of interest to the plastid. The chloroplast transit peptides may be derived from the gene of interest, or may be derived from a heterologous sequence having a CTP. Such transit peptides are known in the art. See, for example, Von Heijne *et al.* (1991) *Plant Mol. Biol. Rep. 9*:104-126; Clark *et al.* (1989) *J. Biol. Chem. 264*:17544-17550; della-Cioppa *et al.* (1987) *Plant Physiol. 84*:965-968; Romer *et al.* (1993) *Biochem. Biophys. Res Commun. 196*:1414-1421; and, Shah *et al.* (1986) *Science 233*:478-481. Additional transit peptides for the translocation of the engineered KAS protein to the endoplasmic reticulum (ER), or vacuole may also find use in the constructs of the present invention.

Depending upon the intended use, additional constructs can be employed containing the nucleic acid sequence which provides for the suppression of the host cell's endogenous KAS protein. Where antisense inhibition of a host cells native KAS protein is desired, the entire wild-type KAS sequence is not required.

The skilled artisan will recognize that there are various methods for the inhibition of expression of endogenous sequences in a host cell. Such methods include, but are not limited to antisense suppression (Smith, *et al.* (1988) *Nature* 334:724-726) , co-suppression (Napoli, *et al.* (1989) *Plant Cell* 2:279-289), ribozymes (PCT Publication WO 97/10328), and combinations of sense and antisense Waterhouse, *et al.* (1998) *Proc. Natl. Acad. Sci. USA* 95:13959-13964. Methods for the suppression of endogenous sequences in a host cell typically employ the transcription or transcription and translation of at least a portion of the sequence to be suppressed. Such sequences may be homologous to coding as well as non-coding regions of the endogenous sequence.

Regulatory transcript termination regions may be provided in plant expression constructs of this invention as well. Transcript termination regions may be provided by the DNA sequence encoding the wild-type KAS or a convenient transcription termination region derived from a different gene source, for example, the transcript termination region which is naturally associated with the transcript initiation region. The skilled artisan will recognize that any convenient transcript termination region which is capable of terminating transcription in a plant cell may be employed in the constructs of the present invention.

Alternatively, constructs may be prepared to direct the expression of the engineered KAS sequences directly from the host plant cell plastid. Such constructs and methods are known in the art and are generally described, for example, in Svab, *et al.* (1990) *Proc. Natl. Acad. Sci. USA* 87:8526-8530 and Svab and Maliga (1993) *Proc. Natl. Acad. Sci. USA* 90:913-917 and in U.S. Patent Number 5,693,507.

A plant cell, tissue, organ, or plant into which the recombinant DNA constructs containing the expression constructs have been introduced is considered transformed, transfected, or transgenic. A transgenic or transformed cell or plant also includes progeny of the cell or plant and progeny produced from a breeding program employing such a transgenic plant as a parent in a cross and exhibiting an altered phenotype resulting from the presence of a engineered KAS nucleic acid sequence.

Plant expression or transcription constructs having an engineered KAS as the DNA sequence of interest for increased or decreased expression thereof may be employed with a wide variety of plant life, particularly, plant life involved in the production of vegetable oils for edible

and industrial uses. Most especially preferred are temperate oilseed crops. Plants of interest include, but are not limited to, rapeseed (Canola and High Erucic Acid varieties), sunflower, safflower, cotton, soybean, peanut, coconut and oil palms, and corn. Depending on the method for introducing the recombinant constructs into the host cell, other DNA sequences may be required. Importantly, this invention is applicable to dicotyledyons and monocotyledons species alike and will be readily applicable to new and/or improved transformation and regulation techniques.

Of particular interest, is the use of engineered KAS constructs in plants which have been genetically engineered to produce a particular fatty acid in the plant seed oil, where TAG in the seeds of nonengineered plants of the engineered species, do not naturally contain that particular fatty acid.

The engineered KAS constructs of the present invention can also be used to provide a means for the production of plants having resistance to plant pathogens. Engineered KAS constructs providing for an increased production of particular fatty acids involved in the biosynthesis of pathogen response signals or inhibitors. For example, engineered KAS constructs providing for the increased production of C:8 fatty acids allows for the production of transgenic plants having an increased tolerance to fungal pathogens.

It is contemplated that the gene sequences may be synthesized, either completely or in part, especially where it is desirable to provide plant-preferred sequences. Thus, all or a portion of the desired structural gene (that portion of the gene which encodes the engineered protein) may be synthesized using codons preferred by a selected host. Host-preferred codons may be determined, for example, from the codons used most frequently in the proteins expressed in a desired host species.

Once the desired engineered KAS nucleic acid sequence is obtained, it may be manipulated in a variety of ways. Where the sequence involves non-coding flanking regions, the flanking regions may be subjected to resection, mutagenesis, etc. Thus, transitions, transversions, deletions, and insertions may be performed on the naturally occurring sequence. In addition, all or part of the sequence may be synthesized. In the structural gene, one or more codons may be modified to provide for a modified amino acid sequence, or one or more codon mutations may be introduced to provide for a convenient restriction site or other purpose

involved with construction or expression. The structural gene may be further modified by employing synthetic adapters, linkers to introduce one or more convenient restriction sites, or the like.

The nucleic acid or amino acid sequences encoding an engineered KAS of this invention may be combined with other non-native, or "heterologous", sequences in a variety of ways. By "heterologous" sequences is meant any sequence which is not naturally found joined to the engineered KAS, including, for example, combinations of nucleic acid sequences from the same plant which are not naturally found joined together.

The DNA sequence encoding an engineered KAS of this invention may be employed in conjunction with all or part of the gene sequences normally associated with the wild-type KAS. In its component parts, a DNA sequence encoding engineered KAS is combined in a DNA construct having, in the 5' to 3' direction of transcription, a transcription initiation control region capable of promoting transcription and translation in a host cell, the DNA sequence encoding engineered KAS and a transcription and translation termination region.

Potential host cells include both prokaryotic and eukaryotic cells. A host cell may be unicellular or found in a multicellular differentiated or undifferentiated organism depending upon the intended use. Cells of this invention may be distinguished by having an engineered KAS foreign to the wild-type cell present therein, for example, by having a recombinant nucleic acid construct encoding an engineered KAS therein.

The methods used for the transformation of the host plant cell are not critical to the present invention. The transformation of the plant is preferably permanent, i.e. by integration of the introduced expression constructs into the host plant genome, so that the introduced constructs are passed onto successive plant generations. The skilled artisan will recognize that a wide variety of transformation techniques exist in the art, and new techniques are continually becoming available. Any technique that is suitable for the target host plant can be employed within the scope of the present invention. For example, the constructs can be introduced in a variety of forms including, but not limited to as a strand of DNA, in a plasmid, or in an artificial chromosome. The introduction of the constructs into the target plant cells can be accomplished by a variety of techniques, including, but not limited to calcium-phosphate-DNA co-precipitation, electroporation, microinjection, *Agrobacterium* infection, liposomes or

microprojectile transformation. The skilled artisan can refer to the literature for details and select suitable techniques for use in the methods of the present invention.

Normally, included with the DNA construct will be a structural gene having the necessary regulatory regions for expression in a host and providing for selection of transformant cells. The gene may provide for resistance to a cytotoxic agent, e.g. antibiotic, heavy metal, toxin, etc., complementation providing prototrophy to an auxotrophic host, viral immunity or the like. Depending upon the number of different host species the expression construct or components thereof are introduced, one or more markers may be employed, where different conditions for selection are used for the different hosts.

Where *Agrobacterium* is used for plant cell transformation, a vector may be used which may be introduced into the *Agrobacterium* host for homologous recombination with T-DNA or the Ti- or Ri-plasmid present in the *Agrobacterium* host. The Ti- or Ri-plasmid containing the T-DNA for recombination may be armed (capable of causing gall formation) or disarmed (incapable of causing gall formation), the latter being permissible, so long as the *vir* genes are present in the transformed *Agrobacterium* host. The armed plasmid can give a mixture of normal plant cells and gall.

In some instances where *Agrobacterium* is used as the vehicle for transforming host plant cells, the expression or transcription construct bordered by the T-DNA border region(s) will be inserted into a broad host range vector capable of replication in *E. coli* and *Agrobacterium*, there being broad host range vectors described in the literature. Commonly used is pRK2 or derivatives thereof. See, for example, Ditta, *et al.*, (*Proc. Nat. Acad. Sci., U.S.A.* (1980) *77*:7347-7351) and EPA 0 120 515, which are incorporated herein by reference. Alternatively, one may insert the sequences to be expressed in plant cells into a vector containing separate replication sequences, one of which stabilizes the vector in *E. coli*, and the other in *Agrobacterium*. See, for example, McBride and Summerfelt (*Plant Mol. Biol.* (1990) *14*:269-276), wherein the pRiHRI (Jouanin, *et al., Mol. Gen. Genet.* (1985) 201:370-374) origin of replication is utilized and provides for added stability of the plant expression vectors in host *Agrobacterium* cells.

Included with the expression construct and the T-DNA will be one or more markers, which allow for selection of transformed Agrobacterium and transformed plant cells. A number

of markers have been developed for use with plant cells, such as resistance to chloramphenicol, kanamycin, the aminoglycoside G418, hygromycin, or the like. The particular marker employed is not essential to this invention, one or another marker being preferred depending on the particular host and the manner of construction.

For transformation of plant cells using *Agrobacterium*, explants may be combined and incubated with the transformed *Agrobacterium* for sufficient time for transformation, the bacteria killed, and the plant cells cultured in an appropriate selective medium. Once callus forms, shoot formation can be encouraged by employing the appropriate plant hormones in accordance with known methods and the shoots transferred to rooting medium for regeneration of plants. The plants may then be grown to seed and the seed used to establish repetitive generations and for isolation of vegetable oils.

There are several possible ways to obtain the plant cells of this invention which contain multiple expression constructs. Any means for producing a plant comprising a construct having a DNA sequence encoding the engineered KAS of the present invention, and at least one other construct having another DNA sequence encoding an enzyme are encompassed by the present invention. For example, the expression construct can be used to transform a plant at the same time as the second construct either by inclusion of both expression constructs in a single transformation vector or by using separate vectors, each of which express desired genes. The second construct can be introduced into a plant which has already been transformed with the engineered KAS expression construct, or alternatively, transformed plants, one expressing the engineered KAS construct and one expressing the second construct, can be crossed to bring the constructs together in the same plant.

Other Constructs and Methods of Use

The invention also relates to vectors that include a polynucleotide or polynucleotides of the invention, host cells that are genetically engineered with vectors of the invention and the production of polypeptides of the invention by recombinant techniques. Cell free translation systems can be employed to produce such protein using RNAs derived from the DNA constructs of the invention.

For recombinant production, host cells can be genetically engineered to incorporate expression systems or portions thereof or polynucleotides of the present invention. Introduction of a polynucleotide into a host cell can be effected by methods described in many standard laboratory manuals, such as Davis et al., Basic Methods in Molecular Biology, (1986) and Sambrook et al, Molecular Cloning: A Laboratory Manual, 2nd Edition, Cold Spring Harbor Laboratory Press, Cold Spring Harbor NY (1989). Such methods include, but are not limited to, calcium phosphate transfection, DEAE dextran mediated transfection, transvection, microinjection, cationic lipid-mediated transfection, electroporation, transduction, scrape loading ballistic introduction and infection.

Representative examples of appropriate hosts include bacterial cells, such as streptococci, staphylococci, enterococci, *E. coli*, streptomyces, and *Bacillus subtilis* cells; fungal cells, such as yeast cells and *Aspergillus* cells; insect cells, such as *Drosophila* S2 and *Spodoptera* Sf9 cells; animal cells such as CHO, COS, HeLa, C127, 3T3, BHK, 293 and Bowes melanoma cells; and plant cells as described above.

A variety of expression systems can be used to produce the polypeptides of the invention. Such vectors include, but are not limited to, chromosomal, episomal, and virus derived vectors, for example vectors from bacterial plasmids, bacteriophage, transposons, yeast episomes, insertion elements, yeast chromosomal elements, viruses such as baculoviruses, papova viruses, such as SB40, vaccinia viruses, adenoviruses, fowl pox viruses, pseudorabies viruses and retroviruses, and vectors derived from combinations of such viruses, such as those derived from plasmid and bacteriophage genetic elements, such as cosmids and phagemids. The expression system constructs may contain control regions that regulate as well as engender expression. Generally, any system or vector which is suitable to maintain, propagate or express polynucleotides and/or to express a polypeptide in a host can be used for expression. The appropriate DNA sequence can be inserted into the chosen expression by any of a variety of well-known and routine techniques, such as, for example, those set forth in Sambrook et al, *Molecular Cloning, A Laboratory Manual*, (*supra*).

Appropriate secretion signals, either homologous or heterologous, can be incorporated into the expressed polypeptide to allow the secretion of the protein into the lumen of the endoplasmic reticulum, the periplasmic space or the extracellular environment.

The polypeptides of the present invention can be recovered and purified from recombinant cell cultures by any of a number of well known methods, including, but not limited to, ammonium sulfate or ethanol precipitation, acid extraction, anion or cation exchange chromatography, phosphocellulose chromatography, hydrophobic interaction chromatography, affinity chromatography, hydroxylapatite chromatography, and lectin chromatography. It is most preferable to use high performance liquid chromatography (HPLC) for purification. Any of the well known techniques for protein refolding can be used to regenerate an active confirmation if the polypeptide is denatured during isolation and/or purification.

The engineered KAS polynucleotides and polypeptides of the present invention find use in a variety of applications.

The engineered KAS polynucleotides and polypeptides as well as the constructs containing such engineered KAS polynucleotides and polypeptides find use in the alteration of fatty acid composition. Furthermore, the engineered KAS polynucleotides and polypeptides of the present invention find use in the production of particular fatty acid components. For example, an engineered KAS having a preference for elongating 6, 8, 10, and 12 carbon acyl-ACP substrates can be used in the production of medium chain fatty acids. Such engineered KAS polynucleotides and polypeptides can also be used with additional sequences for the production of medium chain fatty acids, including, but not limited to, medium chain specific thioesterases (see for example USPN 5,512,482).

The present invention further provides methods for the engineering of polyketides and for the identification of molecules useful in cancer therapy, immunosuppressants, anti-parasite, and antibiotic production.

Thus, the present invention permits the use of molecular design techniques to design, select and synthesize chemical entities and compounds, including inhibitory compounds, capable of binding to the active site or substrate binding site of KAS, in whole or in part.

A first approach enabled by this invention, is to use the structure coordinates of KAS to design compounds that bind to the enzyme and alter the physical properties of the compounds in different ways, e.g., solubility. For example, this invention enables the design of compounds that act as competitive inhibitors of the KAS enzyme by binding to, all or a portion of, the active site of KAS. This invention also enables the design of compounds that act as uncompetitive

inhibitors of the KAS enzyme. These inhibitors may bind to, all or a portion of, the substrate binding site of KAS already bound to its substrate and may be more potent and less non-specific than known competitive inhibitors that compete only for the KAS active site. Similarly, non-competitive inhibitors that bind to and inhibit KAS whether or not it is bound to another chemical entity may be designed using the structure coordinates of KAS of this invention. Additionally, reversible and irreversible inhibitors can also be designed.

A second design approach is to probe KAS with molecules composed of a variety of different chemical entities to determine optimal sites for interaction between candidate ICE inhibitors and the enzyme. For example, high resolution X-ray diffraction data collected from crystals saturated with solvent allows the determination of where each type of solvent molecule sticks. Small molecules that bind tightly to those sites can then be designed and synthesized and tested for their KAS inhibitor activity. Travis, J., *Science*, 262, p. 1374 (1993).

This invention also enables the development of compounds that can isomerize to short-lived reaction intermediates in the chemical reaction of a substrate or other compound that binds to KAS, with KAS. Thus, the time-dependent analysis of structural changes in KAS during its interaction with other molecules is enabled. The reaction intermediates of KAS can also be deduced from the reaction product in co-complex with KAS. Such information is useful to design improved analogues of known KAS inhibitors or to design novel classes of inhibitors based on the reaction intermediates of the KAS enzyme and KAS-inhibitor co-complex. This provides a novel route for designing KAS inhibitors with both high specificity and stability.

Another approach made possible and enabled by this invention, is to screen computationally small molecule data bases for chemical entities or compounds that can bind in whole, or in part, to the KAS enzyme. In this screening, the quality of fit of such entities or compounds to the binding site may be judged either by shape complementarity or by estimated interaction energy. Meng, E. C. *et al.*, *J. Comp. Chem.*, 13, pp. 505-524 (1992).

The invention now being generally described, it will be more readily understood by reference to the following examples which are included for purposes of illustration only and are not intended to limit the present invention.

EXAMPLES

Example 1: Determination of the KAS II-Cerulenin Complex Structure

The KASII-cerulenin complex was prepared as described previously (Edwards, *et al.* (1997) *FEBS Lett.* 402:62-66). Crystals of the complex were grown by the hanging drop method. Droplets consisting of equal amounts of protein solution (6 mg ml^{-1}, 21 protein, 0.3 M NaCl, 25 mM Tris, pH 8.0, 5 mM imidazole, and 10% v/v glycerol) and reservoir solution were equilibrated against 26% w/v polyethylene glycol 8000 and 0.1% v/v 2-mercaptoethanol in water. Data from two crystals were collected at 298 K at the synchrotron in MAX-lab, beamline I711, in Lund. The data was processed with DENZO (Otwinowski (1993) *Proceedings of the Collaborative Computating Project 4 Study Weekend: Data Collection and Processing* (Sawyer, L., Isaacs, N., and Bailey, S.S., eds.) pp 56-62, SERC Daresbury Laboratory, Warrington) and programs from the Collaborative Computating Project 4 Suite (Collaborative Computating Project 4 (1994) *Acta Crystallagr. Sect. D Biol. Crystallogr.* 50:760-763) and the two data sets were scaled together in SCALA (Eavans, (1993) *Proceedings of the Collaborative Computating Project 4 Study Weekend: Data Collection and Processing* (Sawyer, L., Isaacs, N., and Bailey, S.S., eds.) pp 56-62, SERC Daresbury Laboratory, Warrington). The crystals are very radiation-sensitive, but cannot be frozen in a cryostream. Due to non-isomorphism, data of only two crystals could be merged. The crystals of the complex have space group P3221 with similar cell dimensions as the native enzyme. The coordinates of the native enzyme (Huang, *et al.* (1998) *EMBO J.* 17:1183-1191) were used to calculate initial electron density maps with SIGMAA (Read (1986) *Acta Crystallogr.* 42:140-149). All data were used in the refinement; no sigma cutoff was applied. After an initial cycle of positional refinement, the model was rebuilt and a model of cerulenin was included. Further cycles of refinement of the complex were carried out using the program REFMAC (Murshudov, *et al.* (1997) *Acta Crystallagr. Sect. D Biol. Crystallogr* 53:240-253) including a bulk solvent correction, interspersed with inspection and correction of the model using O (Jones, *et al.* (1991) *Acta Crystallagr. Sect. A* 47:100-119), OOPS (Kleywegt, *et al.* (1996) *Acta Crystallagr. Sect. D Biol. Crystallogr* 52:829-832), and

PROCHECK (Laskowski, *et al.* (1993) *J. Appl. Crystallogr.* 26:282-291). Structure comparisons were performed using O (Jones, *et al.* (1991) *supra*) with default parameters.

The complex of KASII from *E. coli* with cerulenin crystallized in space group $P3_121$ isomorphously with the native enzyme (Huang, *et al.*(1998) *supra*), and the crystal structure was determined to 2.65-Å resolution by difference Fourier methods. The final protein model after refinement (*R*-factor 5 0.213 and R_{free} 5 0.270 with good stereochemistry) contains 411 out of the 412 residues of the subunit; no electron density for the N-terminal residue was found. The overall real-space correlation coefficient (Jones, *et al.* (1991) *supra*) is 0.92, and there is well defined electron density for the polypeptide chain except for some side chains on the molecular surface. The inhibitor molecule is well defined by the electron density. However, there is weaker than average electron density for the amide group and no electron density for the last carbon atom of the hydrocarbon tail, indicating considerable flexibility for the terminal methyl group.

The overall structure of the KAS dimer is unchanged upon binding of cerulenin; the root mean square deviations for the 411 Cα atoms of the subunit is 0.23 Å between the two structures. These differences are mainly localized in the active site, in particular in the loop comprising residues 398–401. The main differences in structure between the native enzyme and the cerulenin complex are in the conformation of the side chains of Phe-400 (which was anticipated already from the native structure) and of Ile-108, which have completely new rotamer conformations, and in the positions of the side chains of Cys-163, His-340, and Leu-342, which also have moved substantially. These conformational changes provide access for cerulenin to the active site cysteine and open a hydrophobic pocket for the hydrophobic tail of the inhibitor. From the initial $F_o2\ F_c$ electron density map these structural changes could be readily seen as well as the binding site for the inhibitor). Cerulenin is bound covalently through its C2 carbon atom to the Cys-163 Sγ atom. Its hydrocarbon tail fits in a hydrophobic pocket formed at the dimer interface. The structure of the adduct of cerulenin and cysteine, isolated by tryptic digestion of the cerulenin-fatty acid synthase complex, has been determined by NMR and mass spectroscopy (Funabashi, *et al.* (1989) *J. Biochem.(Tokyo)* 105:751-755). This study revealed that the inhibitor reacts at its C2-epoxide carbon with the SH group of cysteine and that cerulenin formed a hydroxylactam ring. The electron density observed in the KASII-cerulenin complex is not consistent with this structure. It was not possible to model bound cerulenin in the closed ring form but the open form

of the inhibitor could readily be fitted to the electron density map. The hydroxylactam ring, which is formed preferably in protic solvents (Funabashi, *et al.* (1989) *supra*), is not present in the hydrophobic environment of the protein.

In the KASII-cerulenin complex, the inhibitor amide carbonyl oxygen is within hydrogen bond distance to the Nε atoms of the side chains of His-340 and His-303, while the amide NH_2 group does not make any close interactions. It is, however, not possible from the structure to exclude the opposite conformation and interactions for the amide group. The hydroxyl group at C3 forms a hydrogen bond to the main chain NH of Phe-400. The carbonyl oxygen at C4 does not form any polar interactions, in fact, it is located in a very hydrophobic pocket formed by side chains Phe-400, Phe-202, and Val-134 from the other subunit in the dimer. The binding site for the hydrophobic part of the inhibitor is also lined with hydrophobic residues: Ala-162, Gly-107, Leu-342, Phe-202, Leu-111, Ile-108, Ala-193, Gly-198; and from the second subunit in the dimer, Ile-138, Val-134, and Phe-133. The two double bonds with *trans* configuration give the hydrophobic tail a shape that fits to the hydrophobic groove once residue Ile-108 has changed rotamer. In comparison, binding of tetrahydrocerulenin would cost entropy, and as expected it shows more than 2 orders of magnitude less inhibitory activity (D'Agnolo, *et al.*(1973) *Biochim. Biophys. Acta* 326:155-156). The influence of the length of the hydrocarbon chain, maintaining the double bond positions, has been studied using fatty acid synthase from *Saccharomyces cerevisiae* (Morisaki, *et al.* (1993) *J. Biol. Chem.* 211:111-115). Cerulenin (12 carbons) had the highest inhibitory activity, with slightly decreasing binding strength upon increase in chain length. However, when increasing the length from 16 to 18 carbon atoms, the inhibition decreased by 2 orders of magnitude. The size of the hydrophobic pocket in KASII, which binds the hydrocarbon tail of cerulenin, suggests that there is space for a longer hydrophobic tail only if the side chains of Leu-111 and of Phe-133 in the second subunit change their conformation. Thus, possible differences in the sensitivity of condensing enzymes toward cerulenin might be controlled by the size of this cavity.

The structure of the cerulenin complex can be considered to mimic the intermediate formed upon reaction of KAS with the acyl-ACP. In such a complex the hydrophobic cavity would harbor the hydrocarbon tail of the acyl intermediate. The acyl hydrophobic tails will not be restricted by two double bonds (as in the case of cerulenin), and this will allow longer acyl chains

to be buried in this pocket. Inspection of the active site cavity suggests that it would not be possible to harbor a linear acyl chain longer than 14 carbon atoms without structural changes. Such conformational changes must occur since KASII is able to elongate 16:1 to 18:1 (Garwin, *et al.* (1980) *J. Biol. Chem.* 255:3263-3265).

Coordinates for the KAS II crystal structure as well as the KAS-cerulenin complex were produced and are presented in Figures 1 and 2 respectively.

Example 2: Engineering KAS II Proteins

The structure of the *E.coli* KAS II-cerulenin complex was analyzed using the Swiss Pdb Viewer (SPV) modeling program, and by stereo viewing of printouts of the structure in different orientations. Using SPV each of the hydrophobic residues surrounding the bound cerulenin residue were changed to all the possible larger hydrophobic residues, and each of the rotamers for the mutant amino acids were examined for steric clashes (SPV rotamer score) with adjacent amino acids and the bound cerulenin molecule. The identified amino acids were targeted for mutagenesis for decreasing the fatty acid chain length specificity of the KAS II protein. The candidate chain length shortening mutations chosen were those that made the least steric clashes with neighboring amino acids while having the most clashes with the end 1 to 6 carbons of cerulenin.

The structure of the *E.coli* KAS II / cerulenin complex was studied as described above and the hydrophobic amino acid residues near the end of the cerulenin binding "pocket" were identified. These amino acids were identified for mutagenesis for the increase in fatty acid chain length recognition. The large hydrophobic residues positioned beyond the end of the cerulenin potentially preventing longer fatty acids from occupying this pocket were chosen for mutagenesis to smaller (alanine) residues.

PCR site-directed mutagenesis was performed using the Quick-Change™ site-directed mutagenesis kit (Stratagene) following the manufacturers protocol. For the preparation of the specific mutations listed in Table 1, the following oligonucleotide primers were used in the reactions.

Table 1

I108F Sense	5'-GTGCCGCAATTGGATCCGGGTTTGGCGGCCTCGGAC	(SEQ ID NO:1)
Antisense	5'-GTCCGAGGCCGCCAAACCCGGATCCAATTGCGGCAC	(SEQ ID NO:2)
I108L Sense	5'-GTGCCGCAATTGGCTCCGGGCTTGGAGGCCTCGGACTGATCG	(SEQ ID NO:3)
Antisense	5'-CGATCAGTCCGAGGCCTCCAAGCCCGGAGCCAATTGCGGCAC	(SEQ ID NO:4)
A193I Sense	5'-GCAGGTGGCGCCGAGAAAATCAGTACGCCGCTGGGC	(SEQ ID NO:5)
Antisense	5'-GCCCAGCGGCGTACTGATTTTCTCGGCGCCACCTGC	(SEQ ID NO:6)
A193M Sense	5'-GGTGGCGCAGAGAAAATGAGTACTCCGCTGGGCGTTG	(SEQ ID NO:7)
Antisense	5'-CAACGCCCAGCGGAGTACTCATTTTCTCTGCGCCACC	(SEQ ID NO:8)
I108A, L111A, I114A Sense	5'-GCAATTGGCTCCGGGGCTGGCGGCGCCGGACTGGCCGAAG AAAACCACAC	(SEQ ID NO:9)
Antisense	5'-GTGTGGTTTTCTTCGGCCAGTCCGGCGCCGCCAGCCCCGG AGCCAATTGC	(SEQ ID NO:10)
L111A Sense	5'-GGGATTGGCGGCGCCGGACTGATCGAAG	(SEQ ID NO:11)
Antisense	5'-CTTCGATCAGTCCGGCGCCGCCAATCCC	(SEQ ID NO:12)
F133A Sense	5'-GATCAGCCCATTCGCGGTACCGTCAACGATTGTG	(SEQ ID NO:13)
Antisense	5'-CACAATCGTTGACGGTACCGCGAATGGGCTGATC	(SEQ ID NO:14)
I197A Sense	5'-GAGAAAGCCAGTACTCCGGCGGGCGTTGGTGG	(SEQ ID NO:15)
Antisense	5'-CCACCAACGCCCGCCGGAGTACTGGCTTTCTC	(SEQ ID NO:16)

Example 3: Construct Preparation

3A. E. coli Expression Constructs

A series of constructs are prepared to direct the expression of the engineered KAS sequences in *E. coli.*

A series of constructs are prepared to direct the expression of the various engineered KAS sequences in host plant cells.

The construct pCGN10440 contains the I108F mutant expressed from the pQE30 (Qiagen) vector for expression in a host *E. coli* cell.

The construct pCGN10441 contains the I108L mutant expressed from the pQE30 (Qiagen) vector for expression in a host *E. coli* cell.

The construct pCGN10442 contains the A193I mutant expressed from the pQE30 (Qiagen) vector for expression in a host *E. coli* cell.

The construct pCGN10443 contains the I108F, A193I mutant expressed from the pQE30 (Qiagen) vector for expression in a host *E. coli* cell.

The construct pCGN10444 contains the I108L, A193I mutant expressed from the pQE30 (Qiagen) vector for expression in a host *E. coli* cell.

The construct pCGN10445 contains the A193M mutant expressed from the pQE30 (Qiagen) vector for expression in a host *E. coli* cell.

The construct pCGN10446 contains the I108F, A193M mutant expressed from the pQE30 (Qiagen) vector for expression in a host *E. coli* cell.

The construct pCGN10447 contains the I108L, A193M mutant expressed from the pQE30 (Qiagen) vector for expression in a host *E. coli* cell.

The construct pCGN10448 contains the L111A mutant expressed from the pQE30 (Qiagen) vector for expression in a host *E. coli* cell.

The construct pCGN10449 contains the F133A mutant expressed from the pQE30 (Qiagen) vector for expression in a host *E. coli* cell.

The construct pCGN10450 contains the L111A, F133A mutant expressed from the pQE30 (Qiagen) vector for expression in a host *E. coli* cell.

The construct pCGN10451 contains the I108A, L11A, I114A mutant expressed from the pQE30 (Qiagen) vector for expression in a host *E. coli* cell.

The construct pCGN10452 contains the F133A, L197A mutant expressed from the pQE30 (Qiagen) vector for expression in a host *E. coli* cell.

The construct pCGN10453 contains the I108A, L11A, I114A, F133A, L197A mutant expressed from the pQE30 (Qiagen) vector for expression in a host *E. coli* cell.

The construct pCGN10454 contains the L197A mutant expressed from the pQE30 (Qiagen) vector for expression in a host *E. coli* cell.

3B. Preparation of Plant Expression Constructs

A series of constructs are prepared to direct the expression of the engineered KAS sequences in plant host cells, both alone and in combination with additional sequences encoding proteins involved in fatty acid biosynthesis.

A plasmid containing the napin cassette derived from pCGN3223 (described in USPN 5,639,790, the entirety of which is incorporated herein by reference) was modified to make it more useful for cloning large DNA fragments containing multiple restriction sites, and to allow the cloning of multiple napin fusion genes into plant binary transformation vectors. An adapter comprised of the self annealed oligonucleotide of sequence CGCGATTTAAATGGCGCGCCCTGCAGGCGGCCGCCTGCAGGGCGCGCCATTTAAAT (SEQ ID NO:) was ligated into the cloning vector pBC SK+ (Stratagene) after digestion with the restriction endonuclease BssHII to construct vector pCGN7765. Plamids pCGN3223 and pCGN7765 were digested with NotI and ligated together. The resultant vector, pCGN7770, contains the pCGN7765 backbone with the napin seed specific expression cassette from pCGN3223.

A binary vector for plant transformation, pCGN5139, was constructed from pCGN1558 (McBride and Summerfelt, (1990) Plant Molecular Biology, 14:269-276). The polylinker of pCGN1558 was replaced as a HindIII/Asp718 fragment with a polylinker containing unique restriction endonuclease sites, AscI, PacI, XbaI, SwaI, BamHI, and NotI. The Asp718 and HindIII restriction endonuclease sites are retained in pCGN5139.

A binary vector, pCGN8642 was constructed to allow for the rapid cloning of various expression cassettes into the vector for use in plant transformation. The construct contains a multiple cloning region located between the right and left borders of the *Agrobacterium* transfer DNA. The construct also contains the Tn5 gene expressed from the 35S promoter between the multiple cloning site and the left border for selection of transformed plants on kanamycin.

A 354 bp *Bgl*II fragment containing the *Cuphea hookeriana* KASII-7 plastid targeting sequence (Figure 14) (SEQ ID NO:) was cloned into the *Bam*HI site of the various pQE30 constructs containing the *E. coli* KASII (FabF) wild type or mutant KAS sequences. The resultant chimeric KAS II targeting sequence/ FabF encoding sequence were cloned as *Hind*III/ *Sal*I fragments into filled-in *Sal*I/*Xho*I sites of the napin expression cassette, pCGN7770. The resulting napin/KAS cassettes were cloned as *Not*I fragments into the *Not*I sites of various plant binary constructs as described below.

A napin cassette containing the coding sequence of the *Cuphea hookeriana* FatB2 protein (described in PCT Publication WO 98/46776, the entirety of which is incorporated herein by reference) was cloned as a *Not*I fragment into the *Not*I site of pCGN8642 to create pCGN11000.

A napin cassette containing the coding sequence of the *Garm FatA1* protein (described in PCT Publication WO 97/12047, the entirety of which is incorporated herein by reference) was cloned into the *Not*I site of pCGN8642 to create pCGN11003.

A napin cassette containing the native (wild-type) *E. coli* KAS II coding sequence was cloned into the *Not*I site of pCGN11003 to create pCGN11040.

A napin cassette containing the native (wild-type) *E. coli* KAS II coding sequence was cloned into the *Not*I site of pCGN11003 to create pCGN11040.

A napin cassette containing the native (wild-type) *E. coli* KAS II coding sequence was cloned into the *Not*I site of pCGN8642 to create pCGN11041.

A napin cassette containing the native (wild-type) *E. coli* KAS II coding sequence was cloned into the *Not*I site of pCGN11000 to create pCGN11042.

A napin cassette containing the L111A KAS II mutant coding sequence was cloned into the *Not*I site of pCGN11003 to create pCGN11045.

A napin cassette containing the L111A KAS II mutant coding sequence was cloned into the *Not*I site of pCGN8642 to create pCGN11046.

A napin cassette containing the F133A KAS II mutant coding sequence was cloned into the *Not*I site of pCGN11003 to create pCGN11049.

A napin cassette containing the F133A KAS II mutant coding sequence was cloned into the *Not*I site of pCGN11003 to create pCGN11050.

A napin cassette containing the L111A, F133A KAS II double mutant coding sequence was cloned into the *Not*I site of pCGN11003 to create pCGN11053.

A napin cassette containing the L111A, F133A KAS II double mutant coding sequence was cloned into the *Not*I site of pCGN8642 to create pCGN11054.

A napin cassette containing the I108A, L111A, I114A KAS II triple mutant coding sequence was cloned into the *Not*I site of pCGN11003 to create pCGN11057.

A napin cassette containing the I108A, L111A, I114A KAS II triple mutant coding sequence was cloned into the *Not*I site of pCGN8642 to create pCGN11058.

A napin cassette containing the I108A, L111A, I114A, F133A, L197A KAS II multiple mutant coding sequence was cloned into the *Not*I site of pCGN11003 to create pCGN11061.

A napin cassette containing the I108A, L111A, I114A, F133A, L197A KAS IImulitple mutant coding sequence was cloned into the *Not*I site of pCGN8642 to create pCGN11062.

A napin cassette containing the I108F KAS II mutant coding sequence was cloned into the *Not*I site of pCGN11000 to create pCGN11065.

A napin cassette containing the I108F KAS II mutant coding sequence was cloned into the *Not*I site of pCGN8642 to create pCGN11066.

A napin cassette containing the I108F, A193I KAS II double mutant coding sequence was cloned into the *Not*I site of pCGN11000 to create pCGN11069.

A napin cassette containing the I108F, A193I KAS II double mutant coding sequence was cloned into the *Not*I site of pCGN8642 to create pCGN11070.

A napin cassette containing the A193M KAS II mutant coding sequence was cloned into the *Not*I site of pCGN11000 to create pCGN11073.

A napin cassette containing the A193M KAS II mutant coding sequence was cloned into the *Not*I site of pCGN8642 to create pCGN11074.

Example 4: Analysis of Engineered KAS II Proteins Expression in *E. coli*

Figure 7 shows the complete list of mutations that were generated in *E.coli* KAS II using the Stratagene Quick-Change™ site-directed mutagenesis kit, and confirmed by DNA sequencing. The mutant KAS II genes cloned behind an IPTG inducible T5 promoter (pQE30 vector, Qiagen) were transformed into *E.coli* strain M15/pREP4. The effect of the expression of these KAS II mutants on the fatty acid composition of *E.coli* is shown in Figure 3. *E.coli* M15/pREP4 strains containing no vector (-Vec), vector without insert (+Vec), or vectors expression wild-type KAS I or II or single or multiple engineered forms of KASII were grown to mid-log phase in LB media at 30°C. Expression was induced for 2 hours with IPTG (0.75 mM), cells were harvested, lyophilzed, and the lipids were extracted into toluene and derivatized by sodium methoxide and analyzed for fatty acid content by GC FAME analysis as described in Dehesh, *et al.* (1998) *Plant J.* 15:383-390.

The mutations prepared to increase the length of the end product fatty acids lead to the accumulation of abnormally long fatty acids in *E.coli* (Figure 3). Wild-type *E.coli* membranes contain no stearic acid and barely detectable levels of 20:0 and 20:1. Whereas L197, F133A and L111A all resulted in further elongation of the normal membrane components 16:0, and 18:1 resulting in the accumulation of 4, 7 and 13% 18:0 respectively, and 1 to 3% 20:0 and 20:1. KAS II/L111A produced the highest level of 18:0 (13%) while KAS II/L111A-F133A accumulated the highest levels of 20:0 and 20:1 (2 and 4% respectively). Mutations I108A and I114A appeared to decrease the long chain fatty acid accumulation due to L111A and F133A.

The KAS II mutants prepared to shorten the maximum fatty acids were analyzed *in vitro* for the ability to utilize various chain length acyl-ACP substrates. Results of the *in vitro* assays (Figures 4, 5, and 6) demonstrates that the mutants I108F, I108L, A193M, and A193I have a reduced ability to utilize C8-ACP and longer substrates for condensation. However, these mutations are able to utilize C6-ACP substrates for elongation to produce C8 fatty acids. Furthermore, at least one mutation, A193M, had an increased ability to utilize C6-ACP substrates compared to the wild-type KAS for elongation.

The data showing the effect of mutations I108F, I108L, A193I and A193M (together or separately) on the enzymatic activity of KAS II are summarized in figures 4, 5 and 6. Figure 4 shows that mutations I108F, I108L and A193M all cause significant reduction in the activity of KAS II on 8:0-ACP as compared to 6:0-ACP (38, 31 and 12 fold reductions respectively), without significantly reducing the activity on 6:0-ACP. In other words they have effectively changed KAS II into an enzyme capable of making fatty acids up to a maximum of 8 carbons in length. Mutation A193I only causes a 1.8 fold decrease in activity on 8:0-ACP as compared to 6:0-ACP. Figure 5 shows that the combined mutations at I108 and A193 have the effect of reducing the activity of KAS II on 6:0-ACP somewhat, but figure 6 shows that the combined effect was much greater effect on the activity with acyl-ACPs 8:0 and longer (14:0). Consequently the double mutants are even more specific for the synthesis of 8 carbon fatty acids. The most specific is KAS II I108F/A193 KAS II which is 90X more active on 6:0-ACP than it is on 8:0-ACP suggesting that it is now an enzyme highly specific for the synthesis of fatty acids only up to 8 carbons in length.

Example 5: Structural Comparisons of a Plant Medium-Chain specific KAS with *E.coli* KAS II

To further characterize the structure-function relationships of KAS fatty acid binding pockets the modeled structure of a plant medium-chain (8:0, 10:0) specific KAS [*Cuphea. pulcherrima*, (*C.pu*) KASIV] (Dehesh *et al.* (1998) *Plant J.* 15:383-390) was compared with the crystal structure of *E.coli* KAS II. Figure 8 shows that *C.pu* KAS I is predicted to share essentially the same folding pattern as *E.coli* KAS II with the exception of a few loop regions, as might be expected given the structural similarity between KAS enzymes. Furthermore, *Cpu* KAS IV also has a similar structure (Figure 9). The general structure for the KAS family of proteins follows the α-β-α-β-α folding pattern. Indeed at the amino acid sequence level, all but 7 of the 55 highly conserved residues among KAS enzymes are identical (87% identity). However there is only 60% identity in hydrophobic fatty acid binding pocket region with 8 of the 20 amino acids being different consistent with this region of the protein being responsible for the differences in the enzymes specificity. Furthermore the model shows no stearic hinderance in the formation of KASI and KASIV heterodimer (Figure 10). In addition, amino acid sequence comparisons between plant, mammalian, bacterial

Example 6: Plant Transformation and Analysis

The expression constructs described in Example 3B above were used to transform *Arabidopsis thaliana* (Columbia) and/or Columbia mutants *fab*1, *fae*1-1, and *fae*1-2.

Seeds from transformed *Arabidopsis* lines were analyzed for fatty acid composition and are provided in Table 2 below and shown in Figure 13. Fatty acid methyl esters (FAME) extracted in hexane were resolved by gas chromatography (GC) on a Hewlett Packard model 6890 GC.

Table 2

Fatty Acid	12:0	14:0	16:0	16:1	18:0	18:1 C9	18:1 C11	18:2	18:3	20:0	20:1 C11	20:1 C13
11058-AT002-19	0.29	0.17	7.86	0.50	3.85	14.53	3.37	26.02	18.72	2.83	11.61	3.71
11062-AT002-8	0.12	0.00	5.30	0.23	2.49	10.47	1.34	21.55	25.97	2.75	14.55	2.11
AT002-44	0.17	0.00	8.53	0.26	3.35	15.65	1.21	29.06	17.22	2.06	17.22	1.36
11041-AT002-9	0.00	0.00	9.46	0.29	3.49	13.87	1.18	27.32	18.88	2.28	17.52	1.43

Fatty Acid	20:2	20:3	22:0	22:1	22:2	22:3	24:0	24:1
11058-AT002-19	1.39	0.67	0.41	1.71	0.33	0.33	0.90	0.81
11062-AT002-8	2.56	2.07	0.55	5.36	0.40	1.13	0.42	0.63
AT002-44	1.63	0.36	0.29	1.26	0.02	0.07	0.14	0.14
11041-AT002-9	1.69	0.48	0.30	1.46	0.00	0.00	0.18	0.16

T2 pooled seeds from transgenic *Arabidopsis* lines containing pCGN11041 (11041-AT002-9) expressing the native *E. coli* KAS II protein in the seed tissue demonstrated nearly the same fatty acid composition as the nontransformed control *Arabidopsis* plants (AT002-44).

T2 pooled seeds from transgenic *Arabidopsis* var Columbia containing the construct pCGN11058 demonstrated the ability to synthesize longer carbon chain fatty acids compared to the nontransformed control plants as well as transgenic plants containing the wild-type *E. coli* KAS II protein. Particular increases in the production of 18:1 c11, 20:1 c13, 24:0 and 24:1 are observed in transgenic plants containing pCGN11058. Increases of 18:1 c11, 20:1 c13, 24:0 and 24:1 of 2 to 3 fold are obtained compared to nontransformed control plants. The fact that these levels were not higher may be due to the fact that there are many enzymatic steps downstream from the condensation step catalyzed by KAS enzymes which affect the longer chain acyl-ACPs produced incorporation into triglycerides.

T2 pooled seeds from transgenic *Arabidopsis* var Columbia containing the construct pCGN11062 also demonstrated the ability to synthesize longer chain fatty acids compared to nontransformed control plants and transgenic plants containing the wild-type *E. coli* KAS II protein construct. The T2 pooled seeds of 11062 transgenic lines were found to have a 3 to 4 fold increase in 22:1 as well as increased amounts of 20:2, 20:3 and 22:3, consistent with the presence of a KAS II protein being present in the plastid.

The above results demonstrate the ability to modify β-ketoacyl-ACP synthase sequences such that engineered β-ketoacyl-ACP synthases having altered substrate specificity may be produced. Such β-ketoacyl-ACP synthases may be expressed in host cells to provide a supply of the engineered β-ketoacyl-ACP synthase and to modify the existing pathway of fatty acid synthesis such that novel compositions of fatty acids are obtained. In particular, the engineered β-ketoacyl-ACP synthases may be expressed in the seeds of oilseed plants to provide a natural source of desirable TAG molecules.

All publications and patent applications mentioned in this specification are indicative of the level of skill of those skilled in the art to which this invention pertains. All publications and patent applications are herein incorporated by reference to the same extent as if each individual

publication or patent application was specifically and individually indicated to be incorporated by reference.

Although the foregoing invention has been described in some detail by way of illustration and example for purposes of clarity of understanding, it will be obvious that certain changes and modifications may be practiced within the scope of the appended claims.